PCT WELTORGANISATION FÜR GEISTIGES EIGENTUM
Internationales Büro

INTERNATIONALE ANMELDUNG VERÖFFENTLICHT NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT)

(51) Internationale Patentklassifikation 6 :
G06F 15/80

A2

(11) Internationale Veröffentlichungsnummer: **WO 98/34177**

(43) Internationales Veröffentlichungsdatum: 6. August 1998 (06.08.98)

(21) Internationales Aktenzeichen: PCT/DE98/00267

(22) Internationales Anmeldedatum: 29. Januar 1998 (29.01.98)

(30) Prioritätsdaten:
197 03 671.6 31. Januar 1997 (31.01.97) DE

(71) Anmelder *(für alle Bestimmungsstaaten ausser US)*: SIEMENS AKTIENGESELLSCHAFT [DE/DE]; Wittelsbacherplatz 2, D–80333 München (DE).

(72) Erfinder; und
(75) Erfinder/Anmelder *(nur für US)*: TANIGUCHI, Michiaki [JP/DE]; Klenzestrasse 67, D–80469 München (DE). TRESP, Volker [DE/DE]; Kaulbachstrasse 40, D–80539 München (DE).

(81) Bestimmungsstaaten: CN, JP, US, europäisches Patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Veröffentlicht
Ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts.

(54) Title: METHOD FOR COMBINING OUTPUT SIGNALS OF SEVERAL ESTIMATORS, IN PARTICULAR OF AT LEAST ONE NEURAL NETWORK, INTO A RESULTS SIGNAL DETERMINED BY A GLOBAL ESTIMATOR

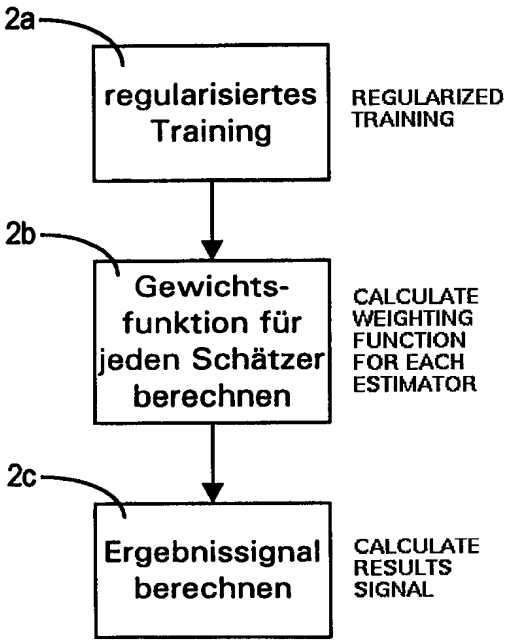
(54) Bezeichnung: VERFAHREN ZUR KOMBINATION VON AUSGANGSSIGNALEN MEHRERER SCHÄTZER, INSBESONDERE VON MINDESTENS EINEM NEURONALEN NETZ, ZU EINEM GESAMTSCHÄTZER ERMITTELTEN ERGEBNISSIGNAL

(57) Abstract

Output signals of individual, computer–assisted statistical estimators (neural networks) are combined into a results signal in a global estimator. To this end, the individual estimators are trained selectively by means of bootstrap data and the weightings of the individual estimators effected in a regularized manner as a contribution to the results signals. Weighting is carried out selectively with "1" or according to the variance of the individual estimator concerned. As a result, significantly better results are achieved for the prediction of numerical values, especially if the amount of training data available is small. Said method can be used for modelling, prognosis and classification by means of neural networks.

(57) Zusammenfassung

Ausgangssignale einzelner rechnergestützter statistischer Schätzer (neuronale Netze) werden zu einem Ergebnissignal in einem Gesamtschätzer zusammengefaßt. Dabei werden die einzelnen Schätzer wahlweise über Bootstrap–Daten trainiert und die Gewichtungen der einzelnen Schätzer als Beitrag zum Ergebnissignal regularisiert vorgenommen. Die Gewichtung erfolgt wahlweise mit "1" oder abhängig von der Varianz des einzelnen Schätzers. Bei der Vorhersage von Zahlenwerten werden somit vor allem bei einer kleinen Menge zur Verfügung stehender Trainingsdaten deutlich bessere Ergebnisse erzielt. Anwendungen: Modellierung, Prognose und Klassifikation mittels neuronaler Netze.



LEDIGLICH ZUR INFORMATION

Codes zur Identifizierung von PCT-Vertragsstaaten auf den Kopfbögen der Schriften, die internationale Anmeldungen gemäss dem PCT veröffentlichen.

AL	Albanien	**ES**	Spanien	**LS**	Lesotho	**SI**	Slowenien
AM	Armenien	**FI**	Finnland	**LT**	Litauen	**SK**	Slowakei
AT	Österreich	**FR**	Frankreich	**LU**	Luxemburg	**SN**	Senegal
AU	Australien	**GA**	Gabun	**LV**	Lettland	**SZ**	Swasiland
AZ	Aserbaidschan	**GB**	Vereinigtes Königreich	**MC**	Monaco	**TD**	Tschad
BA	Bosnien-Herzegowina	**GE**	Georgien	**MD**	Republik Moldau	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagaskar	**TJ**	Tadschikistan
BE	Belgien	**GN**	Guinea	**MK**	Die ehemalige jugoslawische Republik Mazedonien	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Griechenland			**TR**	Türkei
BG	Bulgarien	**HU**	Ungarn	**ML**	Mali	**TT**	Trinidad und Tobago
BJ	Benin	**IE**	Irland	**MN**	Mongolei	**UA**	Ukraine
BR	Brasilien	**IL**	Israel	**MR**	Mauretanien	**UG**	Uganda
BY	Belarus	**IS**	Island	**MW**	Malawi	**US**	Vereinigte Staaten von Amerika
CA	Kanada	**IT**	Italien	**MX**	Mexiko		
CF	Zentralafrikanische Republik	**JP**	Japan	**NE**	Niger	**UZ**	Usbekistan
CG	Kongo	**KE**	Kenia	**NL**	Niederlande	**VN**	Vietnam
CH	Schweiz	**KG**	Kirgisistan	**NO**	Norwegen	**YU**	Jugoslawien
CI	Côte d'Ivoire	**KP**	Demokratische Volksrepublik Korea	**NZ**	Neuseeland	**ZW**	Zimbabwe
CM	Kamerun			**PL**	Polen		
CN	China	**KR**	Republik Korea	**PT**	Portugal		
CU	Kuba	**KZ**	Kasachstan	**RO**	Rumänien		
CZ	Tschechische Republik	**LC**	St. Lucia	**RU**	Russische Föderation		
DE	Deutschland	**LI**	Liechtenstein	**SD**	Sudan		
DK	Dänemark	**LK**	Sri Lanka	**SE**	Schweden		
EE	Estland	**LR**	Liberia	**SG**	Singapur		

Beschreibung

Verfahren zur Kombination von Ausgangssignalen mehrerer Schätzer, insbesondere von mindestens einem neuronalen Netz, zu einem von einem Gesamtschätzer ermittelten Ergebnissignal

Die Erfindung betrifft ein Verfahren zur Kombination von Ausgangssignalen mehrerer Schätzer, insbesondere von mindestens einem neuronalen Netz, zu einem von einem Gesamtschätzer ermittelten Ergebnissignal.

Bei neuronalen Netzen und anderen rechnergestützten Schätzern, die auf statistischen Verfahren beruhen, ist es zweckmäßig, wenn ein Problem einer Zuordnung von Eingangssignalen zu Ausgangssignalen in eine reduzierte Menge von Zuordnungen aufgeteilt wird.

Es ist bekannt, einen rechnergestützten statistischen Schätzer, also beispielsweise ein neuronales Netz, in mehrere kleinere rechnergestützte statistische Schätzer aufzuteilen und Ausgangssignale dieser vereinfachten Schätzer zu gewichten und in einem Gesamtschätzer zu einem Ergebnissignal zu akkumulieren [1].

Um die Ergebnisse bei der Akkumulierung zu verbessern ist aus [2] bekannt, die Ausgangssignale aller Module, also aller einzelnen Schätzer, zu mitteln, und somit das von einem Gesamtschätzer ermittelte Ergebnissignal zu verbessern.

Hierbei sei angemerkt, daß sich der Begriff des statistischen Schätzers nicht ausschließlich auf künstliche neuronale Netze beschränkt, sondern jedwede Art rechnergestützter Schätzer umfaßt, die auf statistischen Verfahren basieren. Eine Übersicht über rechnergestützte Schätzer, die auf statistischen Verfahren basieren ist in [3] gegeben.

Eine Regularisation ist aus [6] bekannt. Insbesondere wird darunter das Hinzufügen eines Terms zu einer Fehlerfunktion verstanden. Ein Regularisationsparameter bestimmt dabei das Gewicht des Terms zur Regularisation.

Aus [7] ist ein Verfahren zur rechnergestützten Kombination einer Vielzahl von Schätzern, die auf statistischen Verfahren beruhen, insbesondere von neuronalen Netzen, zu einem Gesamtschätzer bekannt.

Dem erfindungsgemäßen Verfahren liegt die **Aufgabe** zugrunde, bei einer kleinen Menge von Trainingsdaten ein Überlernen der statistischen Schätzer zu verhindern und somit ein von dem Gesamtschätzer ermitteltes Ergebnissignal zu verbessern.

Die Aufgabe wird gemäß der Merkmale des unabhängigen Patentanspruchs gelöst.

Insbesondere ist für die Erfindung von Bedeutung, daß bei dem Training des einzelnen Schätzers, insbesondere des neuronalen Netzes, eine Regularisation vorgenommen wird. Weiterhin werden die Ausgangssignale der einzelnen Schätzer mit je einer Gewichtsfunktion bewertet. Dabei hängt jede Gewichtsfunktion von den Daten, mit denen der jeweilige Schätzer trainiert wird, ab. Der Gesamtschätzer mittelt die Ausgangssignale aller einzelnen Schätzer zu einem Ergebnissignal. Es ist vorteilhaft, daß bei dem Training jedes einzelnen Schätzers die Kostenfunktion minimiert wird, wobei ein vorgebbarer Regularisationsparameter angibt, inwieweit die Varianz des einzelnen Schätzers verringert wird zu Lasten eines zunehmenden Einflusses von Rauschen. Werden die einzelnen Schätzer mit Bootstrapdaten trainiert, erhält man trotz großer Werte des Regularisationsparameters noch große Varianzen. Unter Bootstrapdaten versteht man Daten, die aus der Menge der Originaldaten durch ziehen mit Zurücklegen gewonnen werden. Für eine nähere Betrachtung der Bootstraptechniken kann auf [4] verwiesen werden.

Eine Weiterbildung des erfindungsgemäßen Verfahrens besteht darin, die Gewichtsfunktion für alle Schätzer zu '1' zu setzen. Weiterhin können die einzelnen Schätzer zusätzlich mit Bootstrapdaten trainiert werden.

Auch kann es vorteilhaft sein, in der Gewichtsfunktion die Varianz zu berücksichtigen. Darüber hinaus können auch in diesem Fall die jeweiligen Schätzer mit Bootstrapdaten trainiert werden.

Weiterbildungen des erfindungsgemäßen Verfahrens ergeben sich aus den abhängigen Ansprüchen.

Die Erfindung wird anhand der folgenden Figuren näher dargestellt.

Es zeigen

Fig.1 ein Blockdiagramm, das schematisch eine Anordnung darstellt, die zeigt, wie die Ausgangssignale mehrerer einzelner Schätzer zu einem Ergebnissignal in einem Gesamtschätzer zusammengeführt werden,

Fig.2 ein Ablaufdiagramm, das die einzelnen Schritte des erfindungsgemäßen Verfahrens darstellt.

In Fig.1 ist dargestellt, wie die Ausgangssignale der einzelnen Schätzer zu einem Ergebnissignal im Gesamtschätzer zusammengeführt werden. Im Folgenden werde angenommen, daß $i=1,2,\ldots,k$ eine Zellvariable für die einzelnen Schätzer ES darstellt. Jeder einzelne Schätzer ES wird durch eine vorgebbare Auswahl von Trainingsdaten x trainiert.

Unter einzelnen Schätzern ES werden in diesem Zusammenhang rechnergestützte statistische Schätzer verstanden, wie beispielsweise neuronale Netze oder jedwede Art anderer Schätzer, die auf statistischen Verfahren beruhen.

Algorithmen zum Trainieren statistischer Schätzer sind dem Fachmann bekannt. Eine Übersicht betreffend eine Auswahl von Trainingsalgorithmen ist in [5] gegeben.

Es ist vorgesehen, daß die einzelnen Schätzer ES nicht alle mit denselben Trainingsdaten x trainiert werden, sondern sie können auch jeweils nur mit einem Teil der Trainingsdaten x trainiert werden, wie das beispielsweise bei Bootstrapdaten der Fall ist.

Durch Gewichtsfunktionen h_i werden Ausgangssignale A_i der einzelnen Schätzer ES gewichtet, wobei die jeweilige Gewichtsfunktion a_i von den Daten, mit denen der jeweilige einzelne Schätzer ES trainiert wird, abhängt.

Das vom Gesamtschätzer GS ermittelte Ergebnissignale ERS liegt am Ausgang des Gesamtschätzer GS an und wird wie folgt berechnet:

$$ERS(x) = \frac{1}{n(x)} \sum_{i=1}^{M} h_i(x) f_i(x) \qquad (1),$$

$$\text{mit} \quad n(x) = \sum_{j=1}^{M} h_j(x) \quad \text{und} \quad h_i(x) \geq 0, \ \forall i = 1,2,..,M,$$

wobei

$h_i(x)$ eine Gewichtsfunktion für den jeweiligen Schätzer und

$n(x)$ einen Normalisierungsfaktor

bezeichnen.

Jeder einzelne Schätzer ES wird regularisiert trainiert, um die Kostenfunktion

$$C_i = \sum_{k=1}^{K} \left(y^k - f_i\left(x^k\right)\right)^2 + \lambda \cdot \sum_{j=1}^{J} w_{ij}^2 \qquad (2)$$

zu minimieren, wobei

$i=1,2,..,M$	den jeweiligen Schätzer,
$K \subseteq L$	eine Menge von K Trainingsdaten aus einer Menge von L Gesamttrainingsdaten,
y	einen Ausgabewert des jeweiligen Schätzers,
$f_i(x)$	eine Antwort des i-ten Schätzers auf die Eingabe x,
$\lambda \geq 0$	einen Regularisationsparameter,
J	eine Anzahl von Gewichten in dem jeweiligen Schätzer und
$\{w_{ij}^2\}_{j=1}^{J}$	Gewichte des i-ten Schätzers,

bezeichnen.

Die Gewichtsfunktion h_i für jeweils alle Schätzer kann gesetzt werden zu

$$h_i(x) = 1 \qquad (3)$$

oder zu

$$h_i(x) = \frac{1}{\mathrm{var}(f_i(x))} \qquad (4),$$

wobei var(fi(x)) die Varianz des Schätzers i für eine Eingabe x bezeichnet.

Es ist vorteilhaft, sowohl die nach Gleichung (3) also auch nach Gleichung (4) gewichteten einzelnen Schätzer jeweils mit Bootstrapdaten zu trainieren.

Anwendung des erfindungsgemäßen Verfahrens ist die Prädiktion von Zahlenfolgen. Eine Menge von vergangenen bekannten Zahlenwerten wird benutzt um einen Schätzer zu trainieren, der entsprechend dem erfindungsgemäßen Verfahren unterteilt ist in einzelne Schätzer und einen Gesamtschätzer. Das somit

trainierte System aus mindestens einem rechnergestützten statistischen Schätzer ist in der Lage, mit einer bestimmten Vorhersagewahrscheinlichkeit den nächsten Wert einer Zahlenfolge vorherzusagen.

Fig.2 zeigt die einzelnen Schritte des erfindungsgemäßen Verfahrens. In Schritt 2a wird ein regularisiertes Training vorgenommen, wobei die Kostenfunktion (2) minimiert werden soll. Dazu muß der Regularisationsparameter λ entsprechend der jeweiligen Anwendung eingestellt werden. Im Schritt 2b wird die Gewichtsfunktion für jeden Schätzer berechnet. Als Gewichtsfunktion wird entweder Formel (3) oder Formel (4) verwendet.

Schließlich wird das Ergebnissignal im Schritt 2c gemäß Formel (1) berechnet.

Im Rahmen dieses Dokuments wurden folgende Veröffentlichungen zitiert:

[1] R. Jacobs et al.: Adaptive Mixtures of Local Experts, Neural Computation, Vol.3, Massachusetts Institute of Technology, 1991, S.79-87.

[2] M. Perrone: Improving Regression Estimates: Averaging Methods for Variance Reduction with Extensions to General Convex Measure Optimization, PhD thesis, Brown University, USA, 1993, S.10-21.

[3] J. Hartung et al.: Statistik, Oldenbourg Verlag, 9.Auflage, München 1993, ISBN 3-486-22055-1, S.123-142.

[4] Efron B. and Tibshirani R.: An Introduction to the Bootstrap, Chapman and Hall 1993, S.45-49.

[5] J. Hertz et al.: Introduction to the Theory of Neural Computation, Addison-Wesley Publishing Company 1991, ISBN 0-201-51560-1, S.89-156.

[6] C. M. Bishop: Neural Networks for Pattern Recognition, Clarendon Press, Oxford 1995, ISBN 0-19 853849 9, Kapitel 5.4, Seiten 171-175.

[7] DE 195 26 954 C1

Patentansprüche

1. Verfahren zur Kombination von Ausgangssignalen mehrerer Schätzer, insbesondere von mindestens einem neuronalen Netz, zu einem von einem Gesamtschätzer ermittelten Ergebnissignal,
 a) bei dem die einzelnen Schätzer als neuronale Netze ausgeführt sind und mit vorgebbaren Daten trainiert werden,
 b) bei dem das Training der Schätzer regularisiert vorgenommen wird,
 c) bei dem das Ausgangssignal jedes einzelnen Schätzers mit einer Gewichtsfunktion bewertet wird,
 d) bei dem jede Gewichtsfunktion von den Daten, mit denen der jeweilige Schätzer trainiert wird, abhängt,
 e) bei dem das Ergebnissignal vom Gesamtschätzer ermittelt wird, indem die Ausgangssignale unter Berücksichtigung der jeweiligen Gewichtsfunktionen gemittelt werden.

2. Verfahren nach Anspruch 1,
bei dem jeder Schätzer regularisiert trainiert wird, um die Kostenfunktion C

$$C_i = \sum_{k=1}^{K} \left(y^k - f_i\left(x^k\right)\right)^2 + \lambda \cdot \sum_{j=1}^{J} w_{ij}^2$$

zu minimieren, wobei

$i=1,2,..,M$	den jeweiligen Schätzer,
$K \subseteq L$	eine Menge von K Trainingsdaten aus einer Menge von L Gesamttrainingsdaten,
y	einen Ausgabewert des jeweiligen Schätzers,
$f_i(x)$	eine Antwort des i-ten Schätzers auf die Eingabe x,
$\lambda \geq 0$	einen Regularisationsparameter,

J eine Anzahl von Gewichten in dem jeweiligen Schätzer,

$\left\{w_{ij}^2\right\}_{j=1}^{J}$ Gewichte des i-ten Schätzers,

bezeichnen.

3. Verfahren nach Anspruch 2,
bei dem sich das Ergebnissignal (ERS) des Gesamtschätzers ergibt aus

$$ERS(x) = \frac{1}{n(x)} \sum_{i=1}^{M} h_i(x) f_i(x)$$

mit $n(x) = \sum_{j=1}^{M} h_j(x)$ und $h_i(x) \geq 0, \forall i = 1,2,..,M,$

wobei

$h_i(x)$ eine Gewichtsfunktion für den jeweiligen Schätzer und

$n(x)$ einen Normalisierungsfaktor,

bezeichnen.

4. Verfahren nach Anspruch 3,
bei dem die Gewichtsfunktion (h) für alle Schätzer gesetzt wird zu

$$h_i(x) = 1, \forall i = 1,2,...,M.$$

5. Verfahren nach Anspruch 4,
bei dem die jeweiligen Schätzer mit Bootstrap-Daten trainiert werden.

6. Verfahren nach Anspruch 3,
bei dem die Gewichtsfunktion (h) für alle Schätzer gesetzt wird zu

$$h_i(x) = \frac{1}{var(f_i(x))}, \forall i = 1,2,\ldots,M,$$

wobei

$var(f_i(x))$ die Varianz des Schätzers i für eine Eingabe x

bezeichnet.

7. Verfahren nach Anspruch 6,
bei dem die jeweiligen Schätzer mit Bootstrap-Daten trainiert werden.

FIG 1

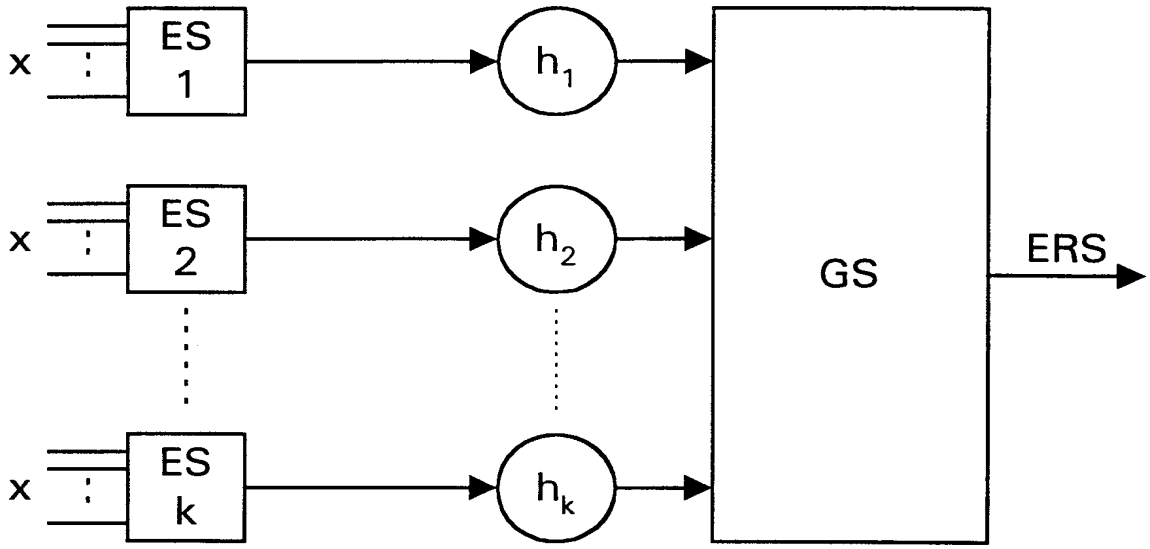
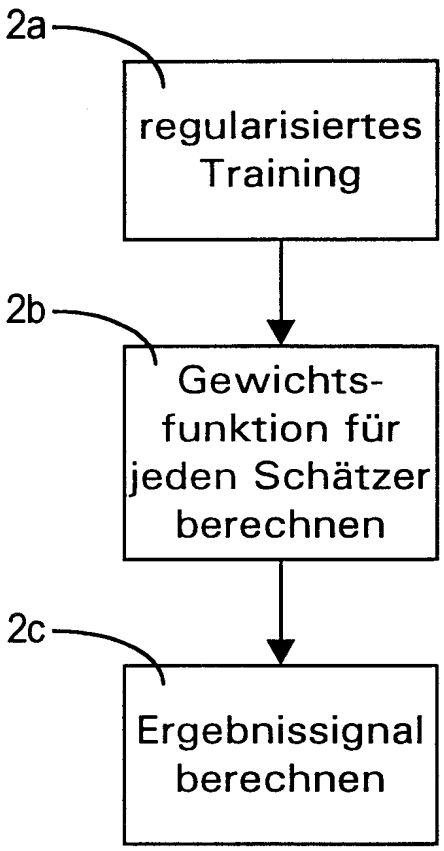


FIG 2


2a
regularisiertes Training
2b
Gewichts-funktion für jeden Schätzer berechnen
2c
Ergebnissignal berechnen